Bayes U: A genomic prediction method based on the Horseshoe prior

Bayes U: A Genomic Prediction Method Based on the Horseshoe Prior

R. Pong-Wong and J.A. Woolliams
The Roslin Institute and R(D)SVS, University of Edinburgh, Midlothian, UK.

ABSTRACT: We propose a novel method for genomic prediction, Bayes U, based on the Horseshoe prior. We compared it with other methods using simulations. All methods compared have different priors for their shrinkage profile. Evaluation of estimated SNP effects showed that Bayes U has stronger variable selection properties, assigning larger estimated effects to those SNPs with strong signals, and assigning more SNPs to have effects closer to zero. However, differences were less noticeable when assessing the accuracy of their overall prediction. Ridge regression and Bayesian Lasso have the lowest accuracies, but no differences were observed with Bayes U, Bayes A, Bayes B and Bayes C. Further studies are required to understand how these methods with different properties lead to similar predictions. The properties of Bayes U may prove to be a desirable behavior for QTL detection and may scale better for sequence data.

Keywords: genomic evaluation; Bayes U; horseshoe prior.

INTRODUCTION

Genomic prediction can be described as the use of high density genotyping in the genetic evaluation to increase the accuracy of the resulting estimated breeding values (GEBV). Several methods have been proposed (e.g. Ridge, Bayes A, Bayes B, Bayes C, Bayesian Lasso) for these predictions. Most use a regression approach where the genotypes for all SNPs are jointly fitted in the model. SNP effects are estimated and thereafter the GEBVs are calculated as the sum of all SNP effects, given the genotypes an individual carries. These methods are defined by the choice of the prior distribution for the SNP effects (P(**β**)) used to prevent problems from over-parameterization due to the large number of SNPs fitted in the model.

In this study we propose a new method of genomic prediction based on the Horseshoe prior and name it as Bayes U. We compare this with five other methods used in genomic prediction and highlight their differences in terms of their shrinkage properties. We compare their predictive properties and accuracy using simulated data.

MATERIALS AND METHODS

Linear model. The basic model is:

$$\mathbf{y} = \mu + \sum \mathbf{z}_i \beta_i + \mathbf{e},$$

where $\mathbf{z}_i$ is the vector of genotype scores at SNP i; βi indicates the allelic substitution effect for SNP i. The prior distributions of SNP effects, P(**β**), defining the Ridge, Bayesian Lasso and Bayes A, are Gaussian, Laplace and scaled Student-t, respectively. The Bayes B and Bayes C have spike and slab priors where a proportion (1-π) of the SNPs have no effect and the remaining (π) with effects distributed as a Student-t and Gaussian, respectively (Nadaf et al. (2012)).

To facilitate the implementation of these methods, P(**β**) is, generally, reformulated as a scale mixture of Normal distributions, by expressing each SNP effect β_i as being distributed $N(0, \sigma_i^2)$ with σ_i^2 randomly sampled from a mixing distribution, $P(\sigma_i^2)$, specific to a given target P(**β**). Hence, $P(\sigma_i^2)$ for Bayesian Lasso is an exponential distribution, and for Bayes A and Bayes B is a scaled inverted χ^2. This hierarchical representation of the model by scale mixtures of Normals allows an easy implementation of the method using Gibbs sampling. Conditional distributions for all parameters required in Ridge, Bayes A, Bayes B, Bayes C and Bayesian Lasso can be found elsewhere (e.g. Gianola et al. (2009)).

Bayes U. The proposed method is based on the Horseshoe prior, which was proposed and described by Carvalho et al. (2010) as having good properties for discriminating between true effect and noise. Assuming this prior, P(**β**) behaves like $\alpha \log(1+\boldsymbol{\beta}^{-2})$ (i.e. up to a constant). It has an infinite spike at zero and heavy tail that decays like β^{-2} (slower than the Laplace or the Student-t).

Similarly, the Horseshoe prior can be reformulated using scale mixtures of Normal distributions, where the mixing distribution is a half Cauchy prior and applied on σ_i (not σ_i^2). Hence, the hierarchical representation of the model is:

$$P(\boldsymbol{\beta}) \propto \log(1+\boldsymbol{\beta}^{-2})$$
$$P(\beta_i) = N(0, \sigma_i^2)$$
$$P(\sigma_i) = C^+(0, \tau)$$

where $C^+(0,\tau)$ is the standard half Cauchy distribution on restricted to $\sigma_i \geq 0$ with scale parameter τ. Carvalho et al. (2010) also proposed the prior distribution of τ to be a half Cauchy prior (i.e. $P(\tau) = C^+(0,\zeta)$). However, to make the results from the Horseshoe prior more comparable with the other methods, a bounded flat prior for τ^2 was used here instead of the half Cauchy on τ. The conditional distribution of σ_i does not have a close form, but sampling this parameter can be done via a slice-sampling approach similar to that proposed by Scott (2011).

A shrinkage parameter. Let k_i be a transformation of σ_i^2 equal to $k_i = 1/(1+\sigma_i^2)$; its distribution has a range between [0,1], with $k_i = 1$ when $\sigma_i^2 = 0$; and $k_i = 0$ when $\sigma_i^2 \to \infty$. Hence, the shrinkage properties of the different methods can be assessed by studying the prior probability distribution function of k_i, $P(k_i)$. The profile close to 0 is associated to the *a priori* weight that the method assigns for *no* shrinkage of the effect (i.e. recognition as a true effect); and the profile close to 1 with the *a priori* weight for full shrinkage (i.e. recognition as noise). Distribution of $P(k_i)$ can be easily obtained by transformation of $P(\sigma_i)$ using standard probability theory. For the case of Bayes U, $P(k_i)$ is a beta distribution with parameters ½ and ½, which has a distinctive U shape.

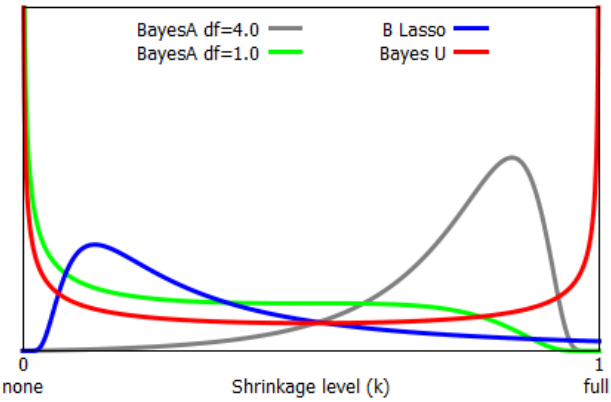

Figure 1. Prior distributions of k for Bayes U, Bayesian Lasso, and Bayes A (with df=1 and 4). Distribution of Bayes B is similar to Bayes A but with a spike on 1 (full shrinkage); Ridge regression is a spike localised at a point related to the variance used; and Bayes C is two spikes at 1 and at a point related to the variance used.

Datasets. The performance of Bayes U was compared with the Ridge regression, Bayesian Lasso, Bayes A, Bayes B and Bayes C using three sets of simulated data.

The first dataset (DATA A) is the smallest and its main purpose is to assess the different methods in terms of the pattern of the estimated effect for each individual SNP included in the analysis. It consisted of 480 phenotyped and genotyped individuals from 6 generations. The simulated trait was assumed to be genetically controlled by 30 unlinked QTL (h^2=0.2) and an extra 480 neutral, unlinked SNPs. Two extra SNPs were also simulated to be in linkage disequilibrium (r^2=0.6) to one of the QTL (QTL1). The analysis was carried assuming that all individuals have known genotypes for the 29 unlinked QTL, the 480 unlinked neutral SNPs and the two SNPs linked to QTL1 (i.e. 511 loci used).

The second and third datasets are from the XV and XVI QTLMAS workshops, denoted as RENNES and SARDINIA respectively. Full description of these datasets can be found in Elsen et al. (2012) and in http://qtl-mas-2012.kassiopeagroup.com/en/index.php. Both sets are divided into a training set with 2000 phenotyped and genotyped individuals, a testing set with 1000 genotyped individuals, and have 10000 SNP. For RENNES, the training and the testing individuals were from the same full/half families. For SARDINIA, a population with three traits over 4 generations was simulated, where the training individuals were from the first three generations and the testing individuals were from the last.

RESULTS AND DISCUSSION

Prior shrinkage profile of the different methods. Figure 1 shows the prior distribution of k for Bayes U, Bayesian Lasso and Bayes A methods. The distribution for Ridge regression is a constant (infinite spike at a location given by the variance used in the model); for Bayes B it is a scaled version of Bayes A plus a spike in 1; and for Bayes C it is two spikes at 1 and with another at a location depending of the variance used. As seen in Figure 1 all methods have very different shrinkage patterns. The main characteristic of Bayes U is that P(k_i) has infinite peaks at k=0 and k=1, implying a strong weight towards applying either full shrinkage on the effects or none at all. On the other hand, P(k_i) for Bayesian Lasso has a value of zero at k=0, implying that all effects will be shrunk to some extent. For the case of Bayes A, the pattern of P(k_i) depends on the assumed degrees of freedom used in P(σ_i^2), and favoring no shrinkage when the degrees of freedom are small.

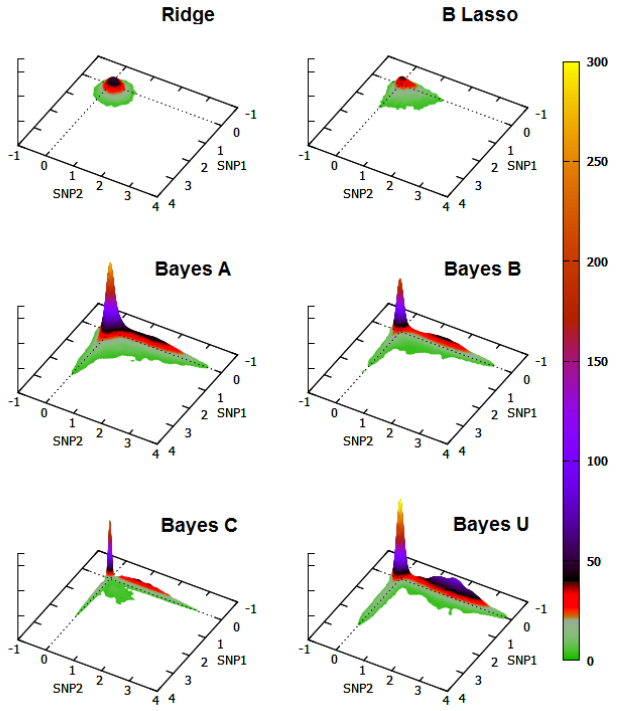

Figure 2. Joint posterior distributions for the effect of SNPs linked to the QTL1 obtained from the different methods of genomic evaluation. X- and Y- axes show the SNP effects, and Z-axes shows density.

The consequences from the differences in the profiles of P(k_i) were clearly reflected in the estimated effects of the SNPs included in the analysis. For instance, Bayes U was the method with the largest SNP effect estimates, and at the same time, the one which assigns more SNPs with (close to) zero effects (results not shown). In other words, the Bayes U was 'sharper' in recognizing (or separating) SNPs with strong and weak 'signals' from the data and applying less shrinkage to the effect of the former but more to the latter. Hence, one may argue that Bayes U has the strongest variable selection behavior by more strongly differentiating what it recognizes as true effects from what it recognizes as noise. After Bayes U the strength of differentiation ranked Bayes B, Bayes A, Bayes C, Bayesian Lasso and Ridge regression. This pattern was consistent across all three datasets.

The differences between the methods can be further observed on the joint posterior distribution for the effect of the 2 SNPs linked to QTL1 from DATA A (Figure 2). The results from Ridge regression suggest that both SNPs have a (small) non-zero effect. However, this is not

Table 1. Accuracy of the overall prediction for the three datsets using the different methods of genomic evaluation.

Method	DATA A	RENNES	SARDINIA		
			Trait1	Trait2	Trait3
Ridge	0.641	0.608	0.738	0.771	0.760
B Lasso	0.663	0.849	0.766	0.809	0.791
Bayes A	0. 690	0.937	0.793	0.834	0.828
Bayes B	0. 697	0.935	0.794	0.833	0.828
Bayes C	0. 698	0.940	0.789	0.820	0.817
Bayes U	0. 697	0.936	0.791	0.825	0.824

the case for Bayes U, Bayes A, Bayes B and Bayes C, where their joint distributions suggest a high probability that neither SNP is affecting the trait, or only one of them (with greater probability on SNP2). Whilst these four methods assign a substantial probability that both SNPs have zero effect, this is larger with Bayes B and Bayes C, reflecting the impact the spike and slap prior assumed in those methods (and this probability is likely to increase if the proportion of SNPs with effect is much lower).

Surprisingly, these noticeable differences between methods in terms of their estimates of SNP effects have little impact on the accuracy of the methods for overall predictions. Table 2 shows the accuracy of the GEBVs obtained for the three datasets. The lowest accuracies were observed with Ridge regression and Bayesian Lasso, but no differences were observed between Bayes A, Bayes B, Bayes C and Bayes U. These results are unexpected considering the diverse behavior of the methods in estimating SNP effects. Possible explanations may range from the size of the data to the genetic models assumed in the simulations. A more comprehensive study with a larger range of genetic models and size of the data is still required to fully understand the overall performance of these methods.

CONCLUSIONS

A novel method for genomic prediction was proposed and its performance compared with other methods commonly used. The proposed method, Bayes U, is based on the Horseshoe prior. We showed that the prior assumed for the SNP effects with these methods have very different shrinkage profiles which affect the behavior of these methods in selecting SNPs for inclusion in the model. These differences, however, were less accentuated when assessing the accuracy of the overall prediction. Further studies are still required to understand the behavior of these methods. However, Bayes U shows desirable selection properties with sharper differentiation between what are recognized as true effects and what is recognized as noise. This may make Bayes U more attractive for QTL detection, and may prove valuable in the analysis of sequence data.

ACKNOWLEDGMENTS

This work was supported by funding from BBSRC Institute Strategic Grant BB/J004235.

LITERATURE CITED

Carvalho, C. M., Polson, N. G., and Scott, J. G. (2010). Biometrika 97:465-480.

Elsen, J. M., Tesseydre, S., Filangi, O. et al. (2012). BMC Proc. 6(Suppl 2):S1.

Gianola, D., de los Campos. G., Hill, W. G. et al. (2009). Genetics 183:347-363.

Nadaf, J., Riggio, V., Yu, T. P. et al. (2012). BMC Proc. 6. (Suppl 2): S6.

Scott, J. G. (2011). Bayesian Anal. 6:307-328.